

# Asymptotic Properties of Order Statistics Correlation Coefficient in the Normal Cases

Weichao Xu, *Member, IEEE*, Chunqi Chang, *Member, IEEE*, Y. S. Hung, *Senior Member, IEEE*, and Peter Chin Wan Fung

**Abstract**—We have previously proposed a novel order statistics correlation coefficient (OSCC), which possesses some desirable advantages when measuring linear and monotone nonlinear associations between two signals. However, the understanding of this new coefficient is far from complete. A lot of theoretical questions, such as the expressions of its distribution and moments, remain to be addressed. Motivated by this unsatisfactory situation, in this paper we prove that for samples drawn from bivariate normal populations, the distribution of OSCC is asymptotically equivalent to that of the Pearson's product moment correlation coefficient (PPMCC). We also reveal its close relationships with the other two coefficients, namely, Gini correlation (GC) and Spearman's rho (SR). Monte Carlo simulation results agree with the theoretical findings.

**Index Terms**—Bivariate normal, concomitant, delta method, Fisher's  $z$  transform, Gini correlation (GC), Kurtosis, order statistics correlation coefficient (OSCC), Pearson's product moment correlation coefficient (PPMCC), ranks, relative efficiency, skewness, Spearman's rho (SR).

## I. INTRODUCTION AND MOTIVATION

A MULTITUDE of methods have been proposed in the literature to measure the intensity of correlation between two random variables with a bivariate distribution. Among these measures the *Pearson's product moment correlation coefficient* [1]–[4] (PPMCC), *Spearman's rho* [5] (SR), and *Kendall's tau* [5] are perhaps the most widely used [6]. The Pearson's coefficient is appropriate mainly for indicating linear associations, while the other two rank-based coefficients are invariant under increasing monotone transformations [5].

Recently, the present authors proposed a novel measure of correlation called *order statistics correlation coefficient* (OSCC), which bridges the gap between Pearson's coefficient and the other two rank-based coefficients [7], [8]. Theoretical analyses and extensive Monte Carlo experiments have shown that OSCC has properties including 1) robustness in the presence of noise, 2) small biasedness, 3) high sensitivity to changes in correlation between signals, 4) capability to detect accurately time-delay, 5) fast computational speed,

and 6) robustness under monotone nonlinear transformations. These desirable properties make OSCC a potentially useful alternative to the three classical correlation coefficients [8]. However, the understanding of OSCC is far from complete due to the lack of knowledge on its distribution with respect to bivariate normal populations. Such knowledge is indispensable when one performs theoretical analyses of OSCC, such as the analytical expressions of its mean, variance, skewness, and kurtosis, just to name a few.

In order to gain further insight into OSCC, we derive in this paper the asymptotic distribution of OSCC when samples are drawn from a bivariate normal population with correlation  $\rho$  (bivariate normal model). We will also compare OSCC to PPMCC, since the latter has long served as a benchmark in the area of correlation studies. In other words, any other correlation coefficients should preferably emulate the properties of PPMCC under the bivariate normal model [9].

The paper is organized as follows. In Section II, we give some basic definitions and several lemmas needed in this paper. In Section III, we prove that OSCC is asymptotically equivalent to PPMCC in terms of distribution and moments. In Section IV, we formulate and prove two theorems on the relationship between OSCC and the other two coefficients. Section V is devoted to the applicability of the asymptotic theories to small samples based on simulation results. Finally, in Section VI, we draw our conclusion on the order statistics correlation coefficient.

## II. PRELIMINARIES

In this section, we give some basic concepts and lemmas concerning normal random variables and the associated order statistics. These prerequisites are necessary in order to establish our main results in the sequel.

### A. Definitions

1) *Order Statistics and Concomitants*: Let  $(X_i, Y_i)$  denote  $n$  independent and identically distributed (i.i.d.) data pairs drawn from a bivariate population with continuous joint cumulative distribution function (cdf). Sorting the data pairs in ascending order with respect to the magnitudes of  $X$ , we get a sequence of new data pairs  $(X_{(i)}, Y_{[i]})$ , where  $X_{(1)} < \dots < X_{(n)}$  are termed the order statistics of  $X$  and  $Y_{[1]}, \dots, Y_{[n]}$  the associated concomitants [10]–[12].

2) *Ranks*: Suppose that  $X_j$  is at the  $k$ th position in the sorted sequence  $X_{(1)} < \dots < X_{(n)}$ , the number  $1 \leq k \leq n$  is termed the rank of  $X_j$  and is denoted by  $p_j$ . Similarly we can get the rank of  $Y_j$  which is denoted by  $q_j$  [5].

Manuscript received March 30, 2007; revised August 25, 2007. This work was supported in part by the Hong Kong Innovation and Technology Commission under Grant ITS/109/02, in part by Hong Kong RGC under Grant N\_HKU703/03, and in part by the University of Hong Kong under Small Project Grant 200507176052. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrice Abry.

W. Xu, C. Chang, and Y. S. Hung are with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong (e-mail: wxu@eee.hku.hk; cqchang@eee.hku.hk; yshung@eee.hku.hk).

P. C. W. Fung is with the Department of Medicine, University of Hong Kong, Hong Kong (e-mail: hrspfcw@hkucc.hku.hk).

Digital Object Identifier 10.1109/TSP.2007.916127

### B. Lemmas

In the sequel, we use symbols  $\mathcal{E}(\cdot)$ ,  $\mathcal{V}(\cdot)$ ,  $\mathcal{C}(\cdot, \cdot)$ , and  $\text{corr}(\cdot, \cdot)$  to denote the mean, variance, covariance, and correlation coefficient of (between) random variables, respectively. We write  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  if  $X$  follows a normal distribution with  $\mathcal{E}(X) = \mu_X$  and  $\mathcal{V}(X) = \sigma_X^2$ , and  $(X, Y) \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$  if  $X$  and  $Y$  follow a bivariate normal distribution with  $\mathcal{E}(X) = \mu_X$ ,  $\mathcal{E}(Y) = \mu_Y$ ,  $\mathcal{V}(X) = \sigma_X^2$ ,  $\mathcal{V}(Y) = \sigma_Y^2$ , and  $\text{corr}(X, Y) = \rho$ . Given these notations, we now list the following lemmas which are necessary to investigate the properties of OSCC under the binormal model.

**Lemma 1:** Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be  $n$  i.i.d. data pairs from a standard bivariate normal population  $\mathcal{N}(0, 0, 1, 1, \rho)$ . Write  $n\bar{X} = \sum X_i$ ,  $n\bar{Y} = \sum Y_i$ ,  $ns_X^2 = \sum (X_i - \bar{X})^2$ ,  $ns_Y^2 = \sum (Y_i - \bar{Y})^2$  and  $Z = s_X - s_Y$ . Then

$$(\bar{X}, \bar{Y}) \sim \mathcal{N}\left(0, 0, \frac{1}{n}, \frac{1}{n}, \rho\right) \quad (1)$$

and

$$Z \sim \mathcal{N}\left(0, \frac{1 - \rho^2}{n}\right) \quad (2)$$

for  $n$  being large.

**Lemma 2:** Let  $\bar{X}, \bar{Y}, s_X, s_Y$  be the same as in Lemma 1. Let  $U_1 = \bar{X}\bar{Y}$ ,  $V_r = s_X s_Y$ , and  $V_1 = (s_X - s_Y)^2 + \bar{X}^2 + \bar{Y}^2$ . Then, as  $n$  large

$$\begin{aligned} \mathcal{E}(U_1) &= O(n^{-1}), \quad \mathcal{V}(U_1) = O(n^{-2}) \\ \mathcal{E}(V_r) &= 1 + O(n^{-1}), \quad \mathcal{V}(V_r) = O(n^{-1}) \\ \mathcal{E}(V_1) &= (n^{-1}), \quad \mathcal{V}(V_1) = O(n^{-2}). \end{aligned} \quad (3)$$

**Lemma 3:** Let  $X_{(1)}, \dots, X_{(n)}$  be the order statistics of an i.i.d. sample  $X_1, \dots, X_n$  drawn from a standard normal population  $X \sim \mathcal{N}(0, 1)$ . Then, as  $n$  large

$$\sum_{i=1}^n \sigma_i^2 = O(\log \log n) \quad (4)$$

where  $\sigma_i^2 = \mathcal{V}(X_{(i)})$  for  $i = 1, \dots, n$ .

**Lemma 4:** Let  $X_{(1)}, \dots, X_{(n)}$  be defined as in Lemma 3. Write  $\mu_i = \mathcal{E}(X_{(i)})$ ,  $\varepsilon_i = X_{(i)} - \mu_i$ ,  $\varepsilon_j = X_{(j)} - \mu_j$ , and  $\sigma_{i,j}^{(2,2)} = \mathcal{E}(\varepsilon_i^2 \varepsilon_j^2)$ . Then, as  $n$  large

$$\sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} = O((\log \log n)^2). \quad (5)$$

The proofs of these lemmas are provided in the Appendices.

### III. ASYMPTOTIC PROPERTIES OF OSCC UNDER BINORMAL MODEL

Let  $(X_i, Y_i)$  be  $n$  i.i.d. data pairs drawn from a bivariate normal population with correlation  $\rho$ . As proposed in [7], [8], the OSCC is defined as

$$w_n(X, Y) \triangleq \frac{\frac{1}{2n} \sum_{i=1}^n [X_{(i)} - X_{(n-i+1)}] Y_{[i]}}{\frac{1}{2n} \sum_{i=1}^n [X_{(i)} - X_{(n-i+1)}] Y_{(i)}}. \quad (6)$$

From its definition in (6), it is easy to verify that OSCC is not symmetric in  $X$  and  $Y$ . However, as pointed out in [8], a symmetric version can be defined as  $w'_n(X, Y) = (1/2)[w_n(X, Y) + w_n(Y, X)]$  if symmetry is a critical feature in practice.

The well-known PPMCC is defined as [2]

$$r_n(X, Y) \triangleq \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (7)$$

It is quite difficult and maybe impossible to derive the exact distribution of  $w_n$  directly from (6). However, as demonstrated below, we can establish the asymptotic equivalence between  $w_n$  and  $r_n$  by writing  $w_n$  as the summation of  $r_n$  and a remainder term, whose mean and variance tend to zero with great rapidity.

#### A. Mean and Variance of $w_n - r_n$ Under Binormal Model

**Theorem 1:** Let  $w_n$  and  $r_n$  be defined as in (6) and (7) with respect to  $n$  i.i.d. sample pairs from a bivariate normal population with correlation  $\rho$ . Write  $e_n = w_n - r_n$ . Then, as  $n \rightarrow \infty$

$$\mathcal{E}(e_n) = O(n^{-1} \log \log n) \quad (8)$$

and

$$\mathcal{V}(e_n) = O(n^{-2} (\log \log n)^2). \quad (9)$$

**Proof:** It has been shown that  $w_n$  is shift and scale invariant [7], [8]. Therefore, without any loss of generality, we assume in the following that the parent population is of the standard bivariate normal distribution, that is

$$(X, Y) \sim \mathcal{N}(0, 0, 1, 1, \rho). \quad (10)$$

Write  $\mathcal{E}(X_{(i)}) = \mu_i$ . We also have  $\mathcal{E}(Y_{(i)}) = \mu_i$  from the assumption (10). Let  $\varepsilon_i = X_{(i)} - \mu_i$  and  $\delta_i = Y_{(i)} - \mu_i$ . It can be shown from [10] that

$$Y_{[i]} = \rho X_{(i)} + z_i \quad (11)$$

where  $X_{(i)}$  are independent of  $z_i$ , the latter being mutually independent with distribution  $\mathcal{N}(0, 1 - \rho^2)$ . Then  $w_n$  can be written as the quotient of [8]

$$U = \frac{1}{n} \sum X_i Y_i - \frac{1}{2n} \sum (\rho \varepsilon_i + z_i)(\varepsilon_i + \varepsilon_{n-i+1}) \quad (12)$$

and

$$V = \frac{1}{2n} \sum (X_i^2 + Y_i^2) - \frac{1}{4n} \left\{ \sum [(\varepsilon_i - \delta_i)^2 + (\varepsilon_i + \delta_{n-i+1})^2] \right\}. \quad (13)$$

Let  $U_r$  and  $V_r$  denote the numerator and denominator of (7), respectively. Let  $\bar{X}, \bar{Y}, s_X$ , and  $s_Y$  be defined as in Lemma 1. After some straightforward algebra, we have

$$U = U_r + U_1 - U_2 \quad (14)$$

and

$$V = V_r + V_1 - V_2 \quad (15)$$

where

$$U_1 = \bar{X}\bar{Y}, \quad (16)$$

$$U_2 = \frac{\rho}{4n} \sum (\varepsilon_i + \varepsilon_{n-i+1})^2 + \frac{1}{2n} \sum z_i (\varepsilon_i + \varepsilon_{n-i+1}) \quad (17)$$

$$V_1 = \frac{1}{2} [(s_X - s_Y)^2 + \bar{X}^2 + \bar{Y}^2] \quad (18)$$

and

$$V_2 = \frac{1}{4n} \left\{ \sum [(\varepsilon_i - \delta_i)^2 + (\varepsilon_i + \delta_{n-i+1})^2] \right\}. \quad (19)$$

Now we can write

$$e_n = w_n - r_n = \frac{U_1 - U_2 - r_n(V_1 - V_2)}{V_r + (V_1 - V_2)}. \quad (20)$$

Let  $U_e$  and  $V_e$  denote the numerator and denominator of (20), respectively. It follows from the *Delta method* [13] that

$$\mathcal{E}(e_n) = \frac{\mathcal{E}(U_e)}{\mathcal{E}(V_e)} + O(n^{-1}) \quad (21)$$

and

$$\mathcal{V}(e_n) \simeq \left[ \frac{\mathcal{E}(U_e)}{\mathcal{E}(V_e)} \right]^2 \left[ \frac{\mathcal{V}(U_e)}{\mathcal{E}^2(U_e)} + \frac{\mathcal{V}(V_e)}{\mathcal{E}^2(V_e)} - \frac{2\mathcal{C}(U_e, V_e)}{\mathcal{E}(U_e)\mathcal{E}(V_e)} \right]. \quad (22)$$

To evaluate (21) and (22), it is sufficient to find  $\mathcal{E}(U_e)$ ,  $\mathcal{E}(V_e)$ ,  $\mathcal{V}(U_e)$ ,  $\mathcal{V}(V_e)$ , and  $\mathcal{C}(U_e, V_e)$ , whose orders of magnitude are determined by  $U_1$ ,  $V_1$ ,  $V_r$ ,  $U_2$ , and  $V_2$ . The asymptotic means and variances of  $U_1$ ,  $V_1$ , and  $V_r$  are provided by Lemma 2. Now we focus on the orders of  $U_2$  and  $V_2$  in the following. Taking expectation of (17) and applying the properties of  $z_i$  as well as some elementary inequalities yield

$$\begin{aligned} 0 \leq \mathcal{E}(U_2) &= \frac{\rho}{4n} \mathcal{E} \left( \sum (\varepsilon_i + \varepsilon_{n-i+1})^2 \right) \\ &\leq \frac{\rho}{n} \mathcal{E} \left( \sum \varepsilon_i^2 \right) = \frac{\rho}{n} \sum \sigma_i^2 = O(n^{-1} \log \log n) \end{aligned} \quad (23)$$

where the last step follows from Lemma 3. By Lemma 3, Lemma 4, and some elementary inequalities, we have

$$\begin{aligned} \mathcal{V}(U_2) &\leq \mathcal{E}(U_2^2) \leq \frac{2\rho^2}{n^2} \sum_{i,j=1}^n \mathcal{E}(\varepsilon_i^2 \varepsilon_j^2) + \frac{2}{n^2} \sum_{i=1}^n \mathcal{E}(z_i^2) \mathcal{E}(\varepsilon_i^2) \\ &= \frac{2\rho^2}{n^2} \sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} + \frac{2(1-\rho^2)}{n^2} \sum_{i=1}^n \sigma_i^2 \\ &= O(n^{-2}(\log \log n)^2) + O(n^{-2} \log \log n) \\ &= O(n^{-2}(\log \log n)^2). \end{aligned} \quad (24)$$

Similarly, we can obtain

$$0 \leq \mathcal{E}(V_2) \leq \frac{1}{n} \sum \mathcal{E}(\varepsilon_i^2) + \frac{1}{n} \sum \mathcal{E}(\delta_i^2) = O(n^{-1} \log \log n) \quad (25)$$

and

$$\begin{aligned} \mathcal{V}(V_2) &\leq \mathcal{E}(V_2^2) \leq \frac{2}{n^2} \sum_{i,j=1}^n \mathcal{E}(\varepsilon_i^2 \varepsilon_j^2 + \delta_i^2 \delta_j^2) \\ &= \frac{4}{n^2} \sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} = O(n^{-2}(\log \log n)^2). \end{aligned} \quad (26)$$

From (7), (18), and (19), it is obvious that  $|r_n| \leq 1$ ,  $V_1 \geq 0$ , and  $V_2 \geq 0$ . Then we have

$$\begin{aligned} -\mathcal{E}(V_1 + V_2) &\leq \mathcal{E}(V_1 - V_2) \leq \mathcal{E}(V_1 + V_2) \\ -\mathcal{E}(V_1 + V_2) &\leq \mathcal{E}(r_n(V_1 - V_2)) \leq \mathcal{E}(V_1 + V_2). \end{aligned} \quad (27)$$

Substituting (3) and (25) into (27) gives

$$\begin{aligned} \mathcal{E}(V_1 - V_2) &= O(n^{-1} \log \log n) \\ \mathcal{E}(r_n(V_1 - V_2)) &= O(n^{-1} \log \log n) \end{aligned} \quad (28)$$

from which and together with (3), (23) we have

$$\begin{aligned} \mathcal{E}(U_e) &= \mathcal{E}(U_1 - U_2) - \mathcal{E}(r_n(V_1 - V_2)) \\ &= O(n^{-1} \log \log n). \end{aligned} \quad (29)$$

By a similar procedure it follows that

$$\mathcal{E}(V_e) = \mathcal{E}(V_r) + \mathcal{E}(V_1 - V_2) = 1 + O(n^{-1} \log \log n). \quad (30)$$

Substituting (29) and (30) into (21) and letting  $n \rightarrow \infty$  yield

$$\mathcal{E}(e_n) = O(n^{-1} \log \log n). \quad (31)$$

By virtual of the Cauchy-Schwarz inequality [14] as well as the results of (3) and (24), it follows that

$$\begin{aligned} 0 \leq \mathcal{V}(U_1 - U_2) &= \mathcal{V}(U_1) + \mathcal{V}(U_2) - 2\mathcal{C}(U_1, U_2) \\ &\leq \left( \sqrt{\mathcal{V}(U_1)} + \sqrt{\mathcal{V}(U_2)} \right)^2 \leq 2[\mathcal{V}(U_1) + \mathcal{V}(U_2)] \\ &= O(n^{-2}) + O(n^{-2}(\log \log n)^2) \\ &= O(n^{-2}(\log \log n)^2) \end{aligned} \quad (32)$$

and similarly

$$\mathcal{V}(V_1 - V_2) = O(n^{-2}(\log \log n)^2). \quad (33)$$

Then we have from (28) and (33) that

$$\begin{aligned} \mathcal{V}(r_n(V_1 - V_2)) &= \mathcal{E}(r_n^2(V_1 - V_2)^2) - \mathcal{E}^2(r_n(V_1 - V_2)) \\ &\leq \mathcal{E}((V_1 - V_2)^2) - \mathcal{E}^2((V_1 - V_2)) \\ &\quad + \mathcal{E}^2((V_1 - V_2)) - \mathcal{E}^2(r_n(V_1 - V_2)) \\ &= \mathcal{V}(V_1 - V_2) + \mathcal{E}^2((V_1 - V_2)) - \mathcal{E}^2(r_n(V_1 - V_2)) \\ &= O(n^{-2}(\log \log n)^2). \end{aligned} \quad (34)$$

Given (3), (32)–(34), and applying the Cauchy-Schwarz inequality again, we have

$$\begin{aligned} \mathcal{V}(U_e) &\leq 2[\mathcal{V}(U_1 - U_2) + \mathcal{V}(r_n(V_1 - V_2))] \\ &= O(n^{-2}(\log \log n)^2) \end{aligned} \quad (35)$$

and

$$\mathcal{V}(V_e) \leq 2[\mathcal{V}(V_r) + \mathcal{V}(V_1 - V_2)] = O(n^{-1}). \quad (36)$$

It follows directly from (35) and (36) that

$$\mathcal{C}(U_e, V_e) = O(n^{-3/2} \log \log n). \quad (37)$$

A substitution of (29), (30), and (35)–(37) into (22) leads to

$$\mathcal{V}(e_n) = O(n^{-2}(\log \log n)^2) \quad \text{as } n \rightarrow \infty \quad (38)$$

whence the result.  $\blacksquare$

*Remark 1:* Since  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  tend to zero as  $n \rightarrow \infty$ , it follows from [15] that  $e_n$  converges in probability to 0, which is denoted by  $e_n \xrightarrow{P} 0$ . On the other hand, it is well known that  $r_n \xrightarrow{P} \rho$ . Then from [16] we have  $w_n \xrightarrow{P} \rho$ , namely,  $w_n$  is an asymptotic consistent estimator of  $\rho$  with respect to the bivariate normal population.

*Remark 2:* Write  $\sigma_e^2 = \mathcal{V}(e_n)$ . It follows directly from (38) that  $\sigma_e = O(n^{-1} \log \log n)$ . By the Chebyshev inequality [13], we have  $P(\|e_n - \mathcal{E}(e_n)\| \leq \lambda \sigma_e) \geq 1 - 1/\lambda^2$  for any positive  $\lambda$ . This shows how the probability of  $e_n$  falling in  $(\mathcal{E}(e_n) - \lambda \sigma_e, \mathcal{E}(e_n) + \lambda \sigma_e)$  can be made arbitrarily close to unity by choosing  $\lambda$  suitably. In other words, we can write

$$e_n = O(n^{-1} \log \log n) \quad (39)$$

as well as

$$w_n = r_n + O(n^{-1} \log \log n). \quad (40)$$

### B. Asymptotic Distribution of OSCC

Having proven Theorem 1 above, we are capable of finding the asymptotic distribution of OSCC based on the relation

$$w_n = r_n + e_n. \quad (41)$$

It can be shown that  $\mathcal{V}(r_n) \simeq (1 - \rho^2)^2 / (n - 1) = O(n^{-1})$  [1]. Then we have

$$\lim_{n \rightarrow \infty} \frac{\mathcal{V}(e_n)}{\mathcal{V}(r_n)} = \lim_{n \rightarrow \infty} \frac{O(n^{-2} (\log \log n)^2)}{O(n^{-1})} = 0. \quad (42)$$

Recall that  $\mathcal{E}(e_n)$  also decreases to zero as  $n$  becomes large. Then the residual term  $e_n$  is negligible for sufficiently large  $n$  [15]. In other words, when  $n$  is large enough, the distribution of  $w_n$  is dominated by the distribution of  $r_n$ , whose density function is, for any  $|w| \leq 1$

$$f_{r_n}(w) = \text{const} (1 - \rho^2)^{(n-1)/2} (1 - w^2)^{(n-4)/2} \frac{\partial^{n-2}}{\partial w^{n-2}} \frac{\theta}{\sin \theta} \quad (43)$$

where  $\theta = \cos^{-1}(-\rho w)$  [2].

### C. Convergence Rate of the Moments of OSCC

*Theorem 2:* Let

$$\mathcal{M}_1(\xi) = \mathcal{E}(\xi) \quad \text{and} \quad \mathcal{M}_k(\xi) = \mathcal{E}([\xi - \mathcal{E}(\xi)]^k)$$

for  $\xi = w_n, r_n$  and  $k \geq 2$ . Then

$$\mathcal{M}_k(w_n) = \mathcal{M}_k(r_n) \left[ 1 + O(n^{-1/2} \log \log n) \right] \quad (44)$$

for  $k = 1, 2, \dots$  and  $n$  being sufficiently large.

*Proof:* For  $k = 1$  and  $n$  large, it follows from [8] and Theorem 1 that

$$\begin{aligned} \mathcal{M}_1(w_n) &= \rho \left[ 1 + O(n^{-1/2} \log \log n) \right] \\ &= \mathcal{M}_1(r_n) \left[ 1 + O(n^{-1/2} \log \log n) \right]. \end{aligned} \quad (45)$$

For  $k \geq 2$ , we have

$$\begin{aligned} \mathcal{M}_k(w_n) &= \sum_{i=0}^k \binom{k}{i} \mathcal{E}((r_n - \rho)^{k-i} e_n^i) \\ &= \mathcal{M}_k(r_n) + k \mathcal{E}((r_n - \rho)^{k-1} e_n) + \dots \\ &\quad + k \mathcal{E}((r_n - \rho) e_n^{k-1}) + \mathcal{M}_k(e_n). \end{aligned} \quad (46)$$

By an argument similar to that given in Remark 2, it follows that

$$r_n - \rho = O(n^{-1/2}) \quad (47)$$

with probability arbitrarily close to unity. The proof is then concluded by substituting (39) and (47) into (46) and neglecting the higher order infinitesimal terms. ■

Given Theorem 2 and the results of Hotelling [17], we can summarize the asymptotic mean, variance, skewness, and kurtosis of  $w_n$  for large  $n$  as

$$\mathcal{E}(w_n) \simeq \rho \quad (48)$$

$$\mathcal{V}(w_n) \simeq \frac{(1 - \rho^2)^2}{n - 1} \quad (49)$$

$$\gamma_1(w_n) \triangleq \frac{\mathcal{M}_3(w_n)}{[\mathcal{M}_2(w_n)]^{3/2}} \simeq \frac{-6\rho}{\sqrt{n-1}} \quad (50)$$

$$\gamma_2(w_n) \triangleq \frac{\mathcal{M}_4(w_n)}{[\mathcal{M}_2(w_n)]^2} - 3 \simeq \frac{6(12\rho^2 - 1)}{n - 1} \quad (51)$$

which are measures of location, scatter, symmetry, and long-tailedness of the distribution [15].

### D. Relative Efficiency of $w_n$ to $r_n$

Since both  $w_n$  and  $r_n$  are unbiased estimators of  $\rho$  for  $n$  large, we can compare the performance of the two estimators by means of the relative efficiency (RE), which can be defined as [18]

$$\text{RE} = \frac{\mathcal{V}(r_n)}{\mathcal{V}(w_n)} \times 100\%. \quad (52)$$

It follows obviously from Theorem 2 that RE approaches 100% as  $n$  becomes sufficiently large.

### E. Fisher's $z$ -Transform

It can be shown that, when  $n$  is large [16]

$$r_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(\rho, \frac{(1 - \rho^2)^2}{n - 1}\right) \quad (53)$$

where the symbol " $\xrightarrow{\mathcal{D}}$ " reads "converges in probability to" [15], [16]. The asymptotic equivalence between  $w_n$  and  $r_n$  established above allows us to assert that

$$w_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(\rho, \frac{(1 - \rho^2)^2}{n - 1}\right) \quad (54)$$

for  $n$  large. However, as pointed out in [13], the use of the relations (53) and (54) is not recommended to test a hypothetical nonzero value of  $\rho$  due to the slow convergence speed of  $r_n$  (and hence  $w_n$ ) to normality and dependence between the standard

error  $(1 - \rho^2)/\sqrt{n-1}$  and  $\rho$ . In one of his pioneering papers [3], Fisher introduced the extremely useful transformation

$$z_r = \frac{1}{2} \log_e \frac{1+r_n}{1-r_n} = \tanh^{-1} r_n \quad (55)$$

and showed that

$$z_r \xrightarrow{D} \mathcal{N}\left(\tanh^{-1} \rho, \frac{1}{n-3}\right) \quad (56)$$

with great rapidity. Write  $z_w = \tanh^{-1} w_n$ . Then we can expect that

$$z_w \xrightarrow{D} \mathcal{N}\left(\tanh^{-1} \rho, \frac{1}{n-3}\right) \quad (57)$$

is also true from the continuous mapping theorem [16].

#### IV. RELATIONS WITH OTHER COEFFICIENTS

It can be shown that OSCC  $w_n(X, Y)$  is closely related to two other coefficients, namely, SR and Gini correlation (GC) [10], [19]. Let  $P = \{p_1, \dots, p_n\}$  and  $Q = \{q_1, \dots, q_n\}$  denote the ranks of  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_n$ , respectively. Then the following two relationships hold.

**Theorem 3:** The order statistics correlation coefficient of the ranks  $w_n(P, Q) = r_S(X, Y)$ , where  $r_S(X, Y)$  denotes Spearman's rho.

**Proof:** It follows obviously that  $p_{(i)} = q_{(i)} = i$  for  $i = 1, \dots, n$ . Substituting these into (6), we have

$$w_n(P, Q) = \frac{\sum (2i - n - 1)q_{[i]}}{\sum (2i - n - 1)i}. \quad (58)$$

Now we evaluate the numerator  $U_3$  and denominator  $V_3$  of (58), respectively, as follows:

$$\begin{aligned} U_3 &= 2 \sum p_i q_i - (n+1) \sum i \\ &= \sum (p_i^2 + q_i^2) - \sum (p_i - q_i)^2 - \frac{n(n+1)^2}{2} \\ &= \frac{n(n^2-1)}{6} - \sum (p_i - q_i)^2 \end{aligned} \quad (59)$$

and

$$V_3 = \sum 2i^2 - (n+1) \sum i = \frac{n(n^2-1)}{6}. \quad (60)$$

Substituting (59) and (60) into (58), we have

$$w_n(P, Q) = 1 - \frac{6 \sum (p_i - q_i)^2}{n(n^2-1)} \quad (61)$$

which is the expression of the Spearman's rho [5]. ■

**Theorem 4:** The OSCC of the ranks of  $X$  and the values of  $Y$  is the sample Gini correlation.

**Proof:** Replacing  $X_i$  by  $p_i$  in (6) yields

$$w_n(P, X) = \frac{\sum (2i - n - 1)Y_{[i]}}{\sum (2i - n - 1)Y_{(i)}} \quad (62)$$

which is the sample Gini correlation  $G(Y, X)$  [19]. ■

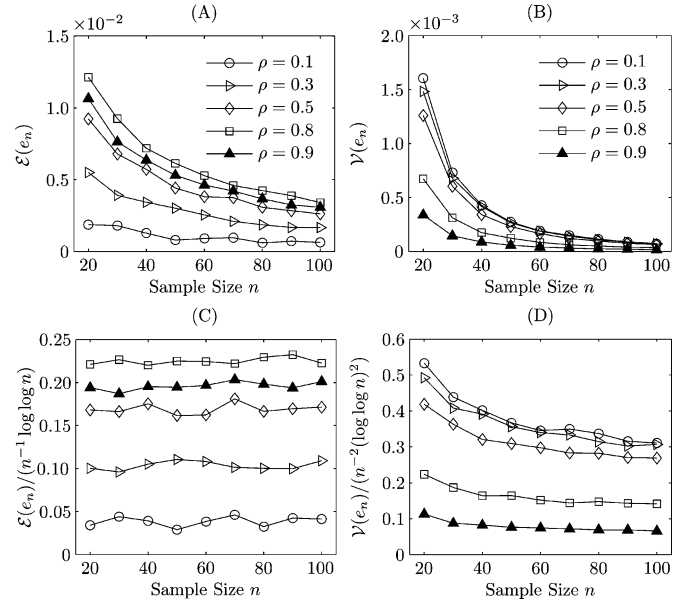


Fig. 1. Illustration of convergence speeds of  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  against  $n = 20, 30, \dots, 100$  and  $\rho = 0.1, 0.3, 0.5, 0.8, 0.9$ , respectively. Note that the legends of (C) and (D), which are the same as those of (A) and (B), are not plotted for the purpose of clarity.

#### V. NUMERIC RESULTS

In this section, we investigate the applicability of our asymptotic theories developed in previous sections for small samples. Monte Carlo experiments are performed for samples of size  $20 \leq n \leq 100$ . The number of trials is set to  $10^5$  for the purpose of accuracy.

##### A. Convergence Rates of $e_n$

Fig. 1 illustrates the convergence speeds of  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  versus the sample size  $n \in [20, 100]$  with an increment  $\Delta n = 10$  in each step. The relationships between the decreasing rates of  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  and the magnitude of  $\rho$  are also revealed by curves associated with  $\rho = 0.1, 0.3, 0.5, 0.8, 0.9$ , respectively. It can be observed that with increase of  $n$ , the magnitudes of  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  decrease downward with quite fast speed. The values of  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  are less than  $10^{-2}$  and  $10^{-3}$ , respectively, for  $n$  even as small as 30. As for the effect of  $\rho$  on the convergence rates,  $\mathcal{E}(e_n)$  and  $\mathcal{V}(e_n)$  behave rather differently. In Fig. 1(A), we can see that for any fixed  $n$ ,  $\mathcal{E}(e_n)$  increases with  $\rho$  at first, but the relation reverses after  $\rho \geq 0.8$ . On the other hand, as shown in Fig. 1(B),  $\mathcal{V}(e_n)$  relates negatively to  $\rho$  in a consistent manner, that is, the larger the intensity of  $\rho$ , the lesser the values of  $\mathcal{V}(e_n)$  for any fixed  $n$ . Fig. 1(C) and (D) depict respectively the ratios  $\aleph_1 = \mathcal{E}(e_n)/(n^{-1} \log \log n)$  and  $\aleph_2 = \mathcal{V}(e_n)/(n^{-2} (\log \log n)^2)$  against  $n \in [20, 100]$  for  $\rho = 0.1, 0.3, 0.5, 0.8, 0.9$ . We can observe that the curves of  $\aleph_1$  are already approximately horizontal when  $n$  is small, suggesting that  $\aleph_1$  converges to constant values with great speed. On the other hand, the convergenc speed of  $\aleph_2$  is rather slow, especially when  $n \leq 50$  and  $\rho \leq 0.5$ . However, despite the magnitude of  $\rho$ , all curves of  $\aleph_2$  become level when  $n \geq 50$ . These observations verify our theoretical results of (8) and (9) established by Theorem 1.

TABLE I  
OBSERVED MEAN VALUES OF  $w_n$  AND  $r_n$  FOR  $n = 20, 50, 100$

$\rho$	$n = 20$		$n = 50$		$n = 100$	
	$w_n$	$r_n$	$w_n$	$r_n$	$w_n$	$r_n$
0.0	0.0020	0.0019	0.0004	0.0004	-0.0003	-0.0003
0.1	0.0981	0.0961	0.0996	0.0986	0.1004	0.0999
0.2	0.1986	0.1950	0.1999	0.1979	0.1998	0.1987
0.3	0.2978	0.2924	0.2997	0.2968	0.2996	0.2980
0.4	0.3978	0.3904	0.4002	0.3964	0.4003	0.3981
0.5	0.4989	0.4899	0.5006	0.4961	0.5007	0.4981
0.6	0.6009	0.5903	0.6019	0.5965	0.6008	0.5977
0.7	0.7020	0.6904	0.7024	0.6964	0.7016	0.6982
0.8	0.8038	0.7919	0.8031	0.7970	0.8019	0.7984
0.9	0.9056	0.8953	0.9035	0.8982	0.9023	0.8993

TABLE II  
COMPARISON RESULTS OF  $\mathcal{V}(w_n)$ ,  $\mathcal{V}(r_n)$  AND THEORETICAL RESULT

$\rho$	$n = 20$			$n = 50$			$n = 100$		
	$w_n$	$r_n$	(49)	$w_n$	$r_n$	(49)	$w_n$	$r_n$	(49)
0.0	0.0564	0.0528	0.0526	0.0211	0.0204	0.0204	0.0103	0.0102	0.0101
0.1	0.0552	0.0517	0.0516	0.0205	0.0198	0.0200	0.0100	0.0098	0.0099
0.2	0.0520	0.0487	0.0485	0.0195	0.0188	0.0188	0.0095	0.0093	0.0093
0.3	0.0476	0.0444	0.0436	0.0176	0.0170	0.0169	0.0085	0.0083	0.0084
0.4	0.0414	0.0386	0.0371	0.0152	0.0147	0.0144	0.0073	0.0072	0.0071
0.5	0.0337	0.0313	0.0296	0.0122	0.0118	0.0115	0.0059	0.0058	0.0057
0.6	0.0253	0.0235	0.0216	0.0090	0.0087	0.0084	0.0043	0.0042	0.0041
0.7	0.0171	0.0159	0.0137	0.0058	0.0056	0.0053	0.0028	0.0027	0.0026
0.8	0.0089	0.0082	0.0068	0.0029	0.0028	0.0026	0.0014	0.0014	0.0013
0.9	0.0026	0.0024	0.0019	0.0008	0.0008	0.0007	0.0004	0.0004	0.0004

### B. Comparative Results for Moments of $w_n$ and $r_n$

Table I lists the observed mean values  $\mathcal{E}(w_n)$  and  $\mathcal{E}(r_n)$  from the Monte Carlo simulations for  $n = 20, 50, 100$ , respectively. It can be seen that 1) the convergence speed of  $\mathcal{E}(w_n)$  to  $\rho$  is quite fast, 2)  $\mathcal{E}(w_n) \geq \mathcal{E}(r_n)$  consistently, 3) for most cases, the biasedness of  $\mathcal{E}(w_n)$  is even smaller than that of  $\mathcal{E}(r_n)$ , and 4) despite few exceptions, the biasedness of  $\mathcal{E}(w_n)$  is negative with small  $\rho$  and positive with large  $\rho$ .

In Table II, we present the simulation results of  $\mathcal{V}(w_n)$  and  $\mathcal{V}(r_n)$  together with theoretical values from (49). Unlike  $\mathcal{E}(w_n)$ , the convergence speeds of  $\mathcal{V}(w_n)$  is rather slow. However, it appears that the difference between  $\mathcal{V}(w_n)$  and  $\mathcal{V}(r_n)$  is less noticeable for  $n = 50$  and negligible for  $n = 100$ . Therefore, it would be safe to approximate  $\mathcal{V}(w_n)$  by (49) for  $n \geq 50$ . A sample size of  $n \geq 100$  can be considered asymptotic in practice.

Tables III and IV contain, respectively, the observed and theoretical values of skewness and kurtosis with respect to  $w_n$  and  $r_n$ . It can be seen that although their convergence speeds to (50) and (51) are much slower,  $\gamma_1(w_n)$  and  $\gamma_2(r_n)$  behave comparably with increase of  $n$  and  $\rho$ . In other words, we do not consider that there are significant differences between  $w_n$  and  $r_n$  as far as convergence rates of their skewness and kurtosis are concerned.

### C. Comparative Results of Relative Efficiency Study

As mentioned in Section III-D, both  $w_n$  and  $r_n$  are eligible as estimators of the population correlation  $\rho$ . Besides, it has been

TABLE III  
OBSERVED AND THEORETICAL VALUES OF  $\gamma_1(w_n)$  AND  $\gamma_1(r_n)$

$\rho$	$n = 20$			$n = 50$			$n = 100$		
	$w_n$	$r_n$	(50)	$w_n$	$r_n$	(50)	$w_n$	$r_n$	(50)
0.0	-0.008	-0.008	0.000	-0.010	-0.012	0.000	-0.001	0.000	0.000
0.1	-0.122	-0.123	-0.138	-0.082	-0.082	-0.086	-0.066	-0.067	-0.060
0.2	-0.243	-0.249	-0.275	-0.167	-0.168	-0.171	-0.107	-0.109	-0.121
0.3	-0.356	-0.364	-0.413	-0.249	-0.251	-0.257	-0.171	-0.170	-0.181
0.4	-0.494	-0.504	-0.551	-0.318	-0.320	-0.343	-0.216	-0.216	-0.241
0.5	-0.637	-0.646	-0.688	-0.414	-0.417	-0.429	-0.291	-0.292	-0.302
0.6	-0.785	-0.786	-0.826	-0.504	-0.506	-0.514	-0.350	-0.351	-0.362
0.7	-0.994	-1.002	-0.964	-0.613	-0.615	-0.600	-0.427	-0.427	-0.422
0.8	-1.187	-1.182	-1.101	-0.695	-0.695	-0.686	-0.497	-0.494	-0.482
0.9	-1.428	-1.396	-1.239	-0.806	-0.792	-0.771	-0.557	-0.551	-0.543

TABLE IV  
OBSERVED AND THEORETICAL VALUES OF  $\gamma_2(w_n)$  AND  $\gamma_2(r_n)$

$\rho$	$n = 20$			$n = 50$			$n = 100$		
	$w_n$	$r_n$	(51)	$w_n$	$r_n$	(51)	$w_n$	$r_n$	(51)
0.0	-0.270	-0.280	-0.316	-0.100	-0.101	-0.122	-0.060	-0.061	-0.061
0.1	-0.239	-0.240	-0.278	-0.091	-0.096	-0.108	-0.058	-0.057	-0.053
0.2	-0.171	-0.172	-0.164	-0.028	-0.029	-0.064	-0.047	-0.044	-0.032
0.3	-0.047	-0.036	0.025	-0.006	-0.006	0.010	-0.002	0.000	0.005
0.4	0.174	0.196	0.291	0.094	0.094	0.113	0.015	0.020	0.056
0.5	0.477	0.491	0.632	0.216	0.219	0.245	0.106	0.103	0.121
0.6	0.842	0.871	1.048	0.370	0.376	0.407	0.195	0.196	0.201
0.7	1.589	1.632	1.541	0.629	0.642	0.598	0.327	0.329	0.296
0.8	2.337	2.348	2.109	0.824	0.838	0.818	0.455	0.449	0.405
0.9	3.668	3.560	2.754	1.150	1.135	1.068	0.553	0.546	0.528

shown that GC and SR can also serve as estimators of  $\rho$ . For convenience, let  $r_s$  and  $r_g$  denote SR and GC, respectively. Then the four estimators are

$$\hat{\rho}_r = r_n \quad (63)$$

$$\hat{\rho}_w = w_n \quad (64)$$

$$\hat{\rho}_g = r_g \quad (65)$$

$$\hat{\rho}_s = 2 \sin \left( \frac{\pi r_s}{6} \right) \quad (66)$$

where  $\hat{\rho}_g$  comes from [19] and  $\hat{\rho}_s$  from [5]. Having (63) – (66), we are able to compute the relative efficiencies of OSCC, GR, and SR (notations  $\text{RE}_w$ ,  $\text{RE}_g$  and  $\text{RE}_s$ ) to Pearson's coefficient  $r_n$  by means of the ratios of  $\mathcal{V}(\hat{\rho}_r)$  to the variance of each of the other three respective estimators.

Fig. 2(A) shows the increasing trend of  $\text{RE}_w$  with respect to sample size  $n \in [20, 100]$ . It also reveals the negative relationship of the convergence speed of  $\text{RE}_w$  to  $\rho$ . It can be observed that despite the negative effect of  $\rho$ , all four curves stand above 96% for  $n \geq 60$  and approach 98% when  $n = 100$ . Fig. 2(B) compares  $\text{RE}_w$ ,  $\text{RE}_g$ , and  $\text{RE}_s$  for  $n \in [20, 100]$  and  $\rho = 0$ . We only compare  $\text{RE}_w$ ,  $\text{RE}_g$  and  $\text{RE}_s$  in the null case due to the lack of theoretical results of  $\text{RE}_s$  except for  $\rho = 0$ . It can be easily seen that  $\text{RE}_w > \text{RE}_g > \text{RE}_s$  for  $n \geq 20$ , suggesting the advantage of OSCC over GC and SR. Moreover, it follows that for  $\rho = 0$ ,  $\text{RE}_g \rightarrow 3/\pi \simeq 95\%$  [20] and  $\text{RE}_s \rightarrow (3/\pi)^2 \simeq 91\%$  [5] as  $n \rightarrow \infty$ . In other words,  $\text{RE}_g$  and  $\text{RE}_s$  can never approach 100% no matter how large  $n$  is. On the other hand,  $\text{RE}_w$  can be made as close to 100% as possible by

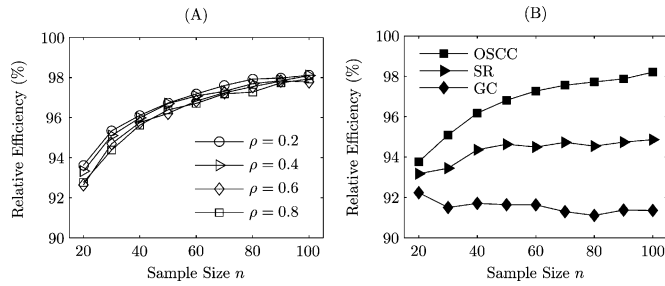


Fig. 2. Relative efficiency of  $w_n$ . (A) relationship of  $RE_w$  versus  $n$  and  $\rho$ . (B) comparison of  $RE_w$ ,  $RE_g$  and  $RE_s$  when  $\rho = 0$

TABLE V  
OBSERVED AND THEORETICAL VALUES OF  $\mathcal{E}(z_w)$  AND  $\mathcal{E}(z_r)$

$\rho$	$\tanh^{-1} \rho$	$n = 20$		$n = 50$		$n = 100$	
		$z_w$	$z_r$	$z_w$	$z_r$	$z_w$	$z_r$
0.0	0.0000	0.0021	0.0020	0.0004	0.0004	-0.0003	-0.0003
0.1	0.1003	0.1043	0.1017	0.1020	0.1010	0.1018	0.1012
0.2	0.2027	0.2132	0.2083	0.2069	0.2047	0.2046	0.2034
0.3	0.3095	0.3253	0.3174	0.3156	0.3122	0.3122	0.3104
0.4	0.4236	0.4454	0.4340	0.4327	0.4277	0.4281	0.4255
0.5	0.5493	0.5787	0.5630	0.5611	0.5545	0.5555	0.5519
0.6	0.6931	0.7326	0.7109	0.7095	0.7002	0.7007	0.6956
0.7	0.8673	0.9179	0.8878	0.8879	0.8749	0.8780	0.8710
0.8	1.0986	1.1650	1.1210	1.1259	1.1071	1.1126	1.1024
0.9	1.4722	1.5725	1.4982	1.5129	1.4816	1.4947	1.4777

TABLE VI  
COMPARISON RESULTS OF  $\mathcal{V}(z_w)$  AND  $\mathcal{V}(z_r)$

$\rho$	$n = 20$		$n = 50$		$n = 100$	
	$z_w$	$z_r$	$z_w$	$z_r$	$z_w$	$z_r$
0.0	0.0636	0.0590	0.0220	0.0213	0.0106	0.0104
0.1	0.0634	0.0588	0.0218	0.0211	0.0104	0.0102
0.2	0.0631	0.0582	0.0220	0.0212	0.0105	0.0103
0.3	0.0638	0.0584	0.0221	0.0212	0.0104	0.0102
0.4	0.0644	0.0583	0.0223	0.0214	0.0105	0.0103
0.5	0.0648	0.0577	0.0223	0.0212	0.0106	0.0103
0.6	0.0661	0.0577	0.0225	0.0212	0.0106	0.0103
0.7	0.0693	0.0583	0.0228	0.0211	0.0107	0.0103
0.8	0.0722	0.0575	0.0234	0.0211	0.0108	0.0103
0.9	0.0814	0.0572	0.0244	0.0210	0.0112	0.0103
$\frac{1}{n-3}$	0.0588		0.0213		0.0103	

choosing  $n$  sufficiently large. This is ensured by the asymptotic equivalence established in Section III.

#### D. Distribution of Fisher's $z$ -Transform

In Table V, we tabulate the expectations of the transformed values  $z_r$ ,  $z_w$  and the asymptotic theoretical expression  $\tanh^{-1} \rho$  corresponding to  $n = 20, 50, 100$ . We can see that  $\mathcal{E}(z_w)$  is consistently greater than  $\mathcal{E}(z_r)$  for all  $\rho \in \{0.0, 0.1, \dots, 0.9\}$ . Furthermore, the biasedness of  $\mathcal{E}(z_w)$  is more noticeable as  $\rho$  large. However, the difference between  $\mathcal{E}(z_w)$  and  $\mathcal{E}(z_r)$  can be considered nonsignificant for  $n \geq 50$ . In Table VI, we summarize the observed variance values  $\mathcal{V}(z_w)$  and  $\mathcal{V}(z_r)$  from the experimental data. For comparison, the theoretical values of  $1/(n-3)$  are also provided in the last row. It can be seen that 1)  $\mathcal{V}(z_w)$  is consistently larger than  $\mathcal{V}(z_r)$ , 2)  $\mathcal{V}(z_w)$  is approximately independent of  $\rho$  for  $n \geq 50$ , and 3) the difference between  $\mathcal{V}(z_w)$  and  $\mathcal{V}(z_r)$  are less noticeable for  $n = 50$  and negligible for  $n = 100$ .

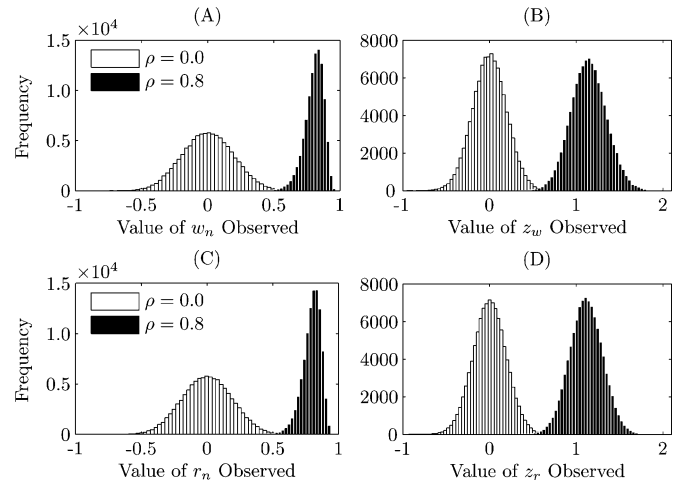


Fig. 3. Empirical distributions of  $w_n$ ,  $z_w$ ,  $r_n$ , and  $z_r$  from populations having correlation 0 and 0.8 when the sample size  $n = 30$ . (A) Histograms of  $w_n$ . (B) Histograms of  $z_w$ . (C) Histograms of  $r_n$ . (D) Histograms of  $z_r$ . It can be observed that the two distributions of  $w_n$  are far from normal for  $n = 30$ . On the contrary, both the two distributions of  $z_w$  are roughly normal with nearly equal variances. Moreover, the distributions of  $w_n$  and  $z_w$  are very similar to their respective counterparts even when  $n$  is as small as 30.

Fig. 3 shows the property of variance-stabilization of Fisher's  $z$  transform. In Fig. 3(A) are plots of the histograms of  $w_n$  ( $n = 30$ ) from populations with correlations 0 and 0.8; Fig. 3(B) shows the corresponding histograms of  $z_w$ . For comparison, the histograms with respect to  $r_n$  and  $z_r$  are also presented in Fig. 3(C) and (D), respectively. The two distributions of  $w_n$  in Fig. 3(A) are drastically distinct in both their modal heights and forms — the one being symmetrical, the other highly skewed. On the other hand, in Fig. 3(B) the two distributions do not differ greatly in height and are approximately the same in form. Besides, the distributions of  $w_n$  and  $z_w$  are very similar to their respective counterparts even when  $n$  is as small as 30. This justifies to some extent the fast convergence rate of  $w_n$  to  $r_n$  in terms of distribution.

#### VI. CONCLUDING REMARKS

In this paper, we have investigated the properties of the order statistics correlation coefficient proposed previously by the present authors. Theoretical derivations and simulation results suggest that the new coefficient is asymptotically equivalent to the Pearson's product moment correlation coefficient in the sense of distribution as well as moments in the normal cases. The new coefficient also has close relationship with the other two correlation coefficients, namely, Spearman's rho and Gini correlation. The advantages of the order statistics correlation coefficient over Pearson's and other correlation coefficients have been discussed in [7] and [8]. The results in this paper further justify that the order statistics correlation coefficient can be used as an alternative to Pearson's coefficient in correlation analysis.

#### APPENDIX I PROOF OF LEMMA 1

*Proof:* It is well known that the normality holds under linear transformations of normal random variables [13]. There-

fore,  $\bar{X}$  and  $\bar{Y}$ , which are linear transformations of normal random variables  $X_i$  and  $Y_i$ , must follow a bivariate normal distribution. Recall that  $(X_i, Y_i) \sim \mathcal{N}(0, 0, 1, 1, \rho)$  and the i.i.d. assumption, we have

$$\begin{aligned}\mathcal{E}(\bar{X}) &= \frac{\sum \mathcal{E}(X_i)}{n} = 0 & \mathcal{E}(\bar{Y}) &= \frac{\sum \mathcal{E}(Y_i)}{n} = 0 \\ \mathcal{V}(\bar{X}) &= \frac{\sum \mathcal{V}(X_i)}{n^2} = \frac{1}{n} & \mathcal{V}(\bar{Y}) &= \frac{\sum \mathcal{V}(Y_i)}{n^2} = \frac{1}{n} \\ \text{corr}(\bar{X}, \bar{Y}) &= \frac{\mathcal{C}(\bar{X}, \bar{Y})}{\sqrt{\mathcal{V}(\bar{X})\mathcal{V}(\bar{Y})}} = \rho\end{aligned}\quad (67)$$

that is,  $(\bar{X}, \bar{Y}) \sim \mathcal{N}(0, 0, 1/n, 1/n, \rho)$ .

It can be shown that  $s_X$  and  $s_Y$  are asymptotically normal [13]. Therefore,  $Z = s_X - s_Y$  is also normally distributed as  $n$  large. From Wishart's formulae [21], we have

$$\mathcal{E}(s_X^2) = \frac{n-1}{n} \quad (68)$$

$$\mathcal{V}(s_X^2) = \frac{2(n-1)}{n^2} \quad (69)$$

and

$$\mathcal{C}(s_X^2, s_Y^2) = \frac{2(n-1)\rho^2}{n^2}. \quad (70)$$

Given (68)–(70) and by the Delta method, we have

$$\mathcal{V}(s_X) \simeq \frac{\mathcal{V}(s_X^2)}{[2\sqrt{\mathcal{E}(s_X^2)}]^2} = \frac{1}{2n} \quad (71)$$

and

$$\mathcal{C}(s_X, s_Y) \simeq \frac{\mathcal{C}(s_X^2, s_Y^2)}{4\sqrt{\mathcal{E}(s_X^2)\mathcal{E}(s_Y^2)}} = \frac{\rho^2}{2n}. \quad (72)$$

Since, by definition,  $\mathcal{V}(s_X) = \mathcal{E}(s_X^2) - \mathcal{E}^2(s_X)$ , then

$$\mathcal{E}(s_X) = [\mathcal{E}(s_X^2) - \mathcal{V}(s_X)]^{1/2} \simeq 1 - \frac{3}{4n}. \quad (73)$$

Noticing that  $\mathcal{E}(s_X) = \mathcal{E}(s_Y)$  and  $\mathcal{V}(s_X) = \mathcal{V}(s_Y)$ , we finally arrive at

$$\mathcal{E}(Z) = 0 \quad (74)$$

and

$$\mathcal{V}(Z) = \mathcal{V}(s_X) + \mathcal{V}(s_Y) - 2\mathcal{C}(s_X, s_Y) = \frac{1-\rho^2}{n} \quad (75)$$

thus establishing the lemma.  $\blacksquare$

## APPENDIX II PROOF OF LEMMA 2

*Proof:* From (1) in Lemma 1 we have

$$\mathcal{E}(U_1) = \mathcal{E}(\bar{X}\bar{Y}) = \mathcal{C}(\bar{X}, \bar{Y}) = \frac{\rho}{n} = O(n^{-1}) \quad (76)$$

and

$$\begin{aligned}\mathcal{V}(U_1) &= \mathcal{V}(\bar{X}\bar{Y}) = \mathcal{E}(\bar{X}^2\bar{Y}^2) - \mathcal{E}^2(\bar{X}\bar{Y}) \\ &= \frac{1+2\rho^2}{n^2} - \frac{\rho^2}{n^2} = \frac{1+\rho^2}{n^2} = O(n^{-2}).\end{aligned}\quad (77)$$

From (2) it follows that

$$\begin{aligned}\mathcal{E}(V_r) &= \mathcal{E}(s_X s_Y) = \mathcal{C}(s_X, s_Y) + \mathcal{E}(s_X)\mathcal{E}(s_Y) \\ &\simeq \frac{\rho^2}{2n} + \left(1 - \frac{3}{4n}\right)^2 = 1 + O(n^{-1}).\end{aligned}\quad (78)$$

Applying some basic identities together with (68), (70), and (78), we have

$$\begin{aligned}\mathcal{V}(V_r) &= \mathcal{V}(s_X s_Y) = \mathcal{E}(s_X^2 s_Y^2) - \mathcal{E}^2(s_X s_Y) \\ &= \mathcal{C}(s_X^2, s_Y^2) + \mathcal{E}(s_X^2)\mathcal{E}(s_Y^2) - \mathcal{E}^2(s_X s_Y) \\ &= \frac{1+\rho^2}{n} + O(n^{-2}) = O(n^{-1}).\end{aligned}\quad (79)$$

For brevity, let  $W = \bar{X}^2 + \bar{Y}^2$ . It has been known long ago that  $(s_X, s_Y)$  is independent of  $(\bar{X}, \bar{Y})$  [13]. Hence

$$\begin{aligned}\mathcal{E}(V_1) &= \mathcal{E}(Z^2 + W) = \mathcal{E}(Z^2) + \mathcal{E}(W) \\ \mathcal{V}(V_1) &= \mathcal{V}(Z^2 + W) = \mathcal{V}(Z^2) + \mathcal{V}(W).\end{aligned}\quad (80)$$

We have shown in Lemma 1 that  $Z \sim \mathcal{N}(0, (1-\rho^2)/n)$  as  $n$  large. Then we have

$$\begin{aligned}\mathcal{E}(Z^2) &= \mathcal{V}(Z) \simeq \frac{1-\rho^2}{n} \\ \mathcal{V}(Z^2) &= \mathcal{E}(Z^4) - \mathcal{E}^2(Z^2) \simeq \frac{2(1-\rho^2)^2}{n^2}.\end{aligned}\quad (81)$$

On the other hand, from (1) it follows that

$$\begin{aligned}\mathcal{E}(W) &= \mathcal{E}(\bar{X}^2) + \mathcal{E}(\bar{Y}^2) = \frac{2}{n} \\ \mathcal{V}(W) &= \mathcal{V}(\bar{X}^2) + \mathcal{V}(\bar{Y}^2) + 2\mathcal{C}(\bar{X}^2, \bar{Y}^2) \\ &= \mathcal{V}(\bar{X}^2) + \mathcal{V}(\bar{Y}^2) + 2[\mathcal{E}(\bar{X}^2\bar{Y}^2) - \mathcal{E}(\bar{X}^2)\mathcal{E}(\bar{Y}^2)] \\ &= \frac{2}{n^2} + \frac{2}{n^2} + \frac{4\rho^2}{n^2} = \frac{4(1+\rho^2)}{n^2}.\end{aligned}\quad (82)$$

Substituting (81) and (82) into (80) leads to

$$\begin{aligned}\mathcal{E}(V_1) &\simeq \frac{3-\rho^2}{n} = O(n^{-1}) \\ \mathcal{V}(V_1) &\simeq \frac{2(3+\rho^4)}{n^2} = O(n^{-2})\end{aligned}\quad (83)$$

thus completes the proof.  $\blacksquare$

## APPENDIX III PROOF OF LEMMA 3

*Proof:* Write  $u_i = i/(n+1)$  and  $v_i = 1 - u_i$ . Let  $\Phi(x)$  be the cdf of  $X$  and  $W(\cdot) = \Phi^{-1}(\cdot)$ . It follows [10] that, as

$$n \rightarrow \infty \quad \sum_{i=1}^n \sigma_i^2 = \left\{ \frac{1}{n+2} \sum_{i=1}^n u_i v_i [W'(u_i)]^2 \right\} = J \quad (\text{say}). \quad (84)$$



It can be easily shown that  $W'(x) = 1/\phi(\Phi^{-1}(x))$ , with  $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  being the pdf of  $X$  [10]. Writing  $\Delta = 1/(n+1)$  and substituting the expression of  $W'$  into (84) yield

$$J < \sum_{i=1}^n \frac{u_i(1-u_i)}{\phi^2(\Phi^{-1}(u_i))} \Delta = \sum \Xi(u_i) \Delta \quad (\text{say}). \quad (85)$$

We can prove that the function  $\Xi(u)$  is nondecreasing over  $u \in [1/2, 1)$  by verifying the nonnegativity of its derivative, as

$$\Xi'(u) = \frac{(1-2u)\phi^2(\Phi_u^{-1}) + 2u(1-u)\phi(\Phi_u^{-1})\Phi_u^{-1}}{\phi^4(\Phi_u^{-1})} \geq 0 \quad (86)$$

where  $\Phi_u^{-1}$  is a compact notation of  $\Phi^{-1}(u)$ . Write  $I = \int_{1/2}^{1-\Delta/2} \Xi(u) du$ . Now we show that when  $n$  is fix,  $J < 4I$ . Noticing that  $\Xi(u_i) = \Xi(u_{n-i+1})$ , we have, for  $n$  odd

$$\begin{aligned} \sum_{i=1}^n \Xi(u_i) \Delta &= \Xi\left(\frac{1}{2}\right) \Delta + 2 \sum_{i=(n+3)/2}^n \Xi(u_i) \Delta \\ &< 4 \left[ \Xi\left(\frac{1}{2}\right) + \cdots + \Xi(u_{n-1}) \right] \Delta \\ &+ 4\Xi(u_n) \frac{\Delta}{2} = 4S_1 \quad (\text{say}) \end{aligned} \quad (87)$$

and for  $n$  even

$$\begin{aligned} \sum_{i=1}^n \Xi(u_i) \Delta &= 2 \sum_{i=(n+2)/2}^n \Xi(u_i) \Delta \\ &< 4 \left[ \Xi(u_{(n+2)/2}) + \cdots + \Xi(u_{n-1}) \right] \Delta \\ &+ 4 \left[ \Xi\left(\frac{1}{2}\right) + \Xi(u_n) \right] \frac{\Delta}{2} \\ &= 4S_2 \quad (\text{say}). \end{aligned} \quad (88)$$

It is easy to verify that both  $S_1$  and  $S_2$  are left Riemann sums of  $\Xi(u)$  over  $u \in [1/2, 1 - \Delta/2]$ . Therefore,  $S_1$  and  $S_2$  are underestimations of  $I$  since  $\Xi(u)$  is nondecreasing over  $u \in [1/2, 1)$ , that is,  $J < 4I$  for any fix  $n$ . Write  $t = \Phi^{-1}(x)$ . We have

$$I = \int_0^{\Phi^{-1}(1-\Delta/2)} \frac{\Phi(t)[1-\Phi(t)]}{\phi(t)} dt. \quad (89)$$

Let  $\zeta = \sqrt{2 \log(n+1)}$ . Now we show that  $\Phi^{-1}(1 - \Delta/2) < \zeta$ . We first prove that  $\Phi^2(-\zeta) < \Delta^2/4$ , as

$$\begin{aligned} \Phi^2(-\zeta) &= \frac{1}{2\pi} \int_{\zeta}^{\infty} \int_{\zeta}^{\infty} e^{-(x^2+y^2)/2} dx dy \\ &< \frac{1}{2\pi} \int_0^{\pi/2} d\theta \int_{\sqrt{2}\zeta}^{\infty} e^{-r^2/2} r dr = \frac{1}{4} \frac{1}{(n+1)^2}. \end{aligned} \quad (90)$$

Then we have  $\Phi(\zeta) = 1 - \Phi(-\zeta) > 1 - \Delta/2$  and hence

$$\Phi^{-1}\left(1 - \frac{\Delta}{2}\right) < \zeta. \quad (91)$$

Substituting (91) into (89) leads to

$$I < \int_0^{\zeta} \frac{\Phi(t)[1-\Phi(t)]}{\phi(t)} dt. \quad (92)$$

Write  $\Gamma(\zeta) = \Phi(\zeta)[1-\Phi(\zeta)]/\phi(\zeta)$ . Then it follows directly from the L'Hospital's Rule that

$$\Gamma(\zeta) \sim \zeta^{-1} \quad \text{as } \zeta \rightarrow \infty. \quad (93)$$

Applying the L'Hospital's Rule once more, we have

$$\lim_{\zeta \rightarrow \infty} \frac{\int_0^{\zeta} \Gamma(t) dt}{\log \zeta} = \lim_{\zeta \rightarrow \infty} \frac{\left[ \int_0^{\zeta} \Gamma(t) dt \right]'}{(\log \zeta)'} = \lim_{\zeta \rightarrow \infty} \frac{\Gamma(\zeta)}{\zeta^{-1}} = 1. \quad (94)$$

Then, for  $n$  being sufficiently large

$$\sum_{i=1}^n \sigma_i^2 < 4 \int_0^{\zeta} \Gamma(t) dt = O(\log \zeta) = O(\log \log n) \quad (95)$$

whence the result. ■

#### APPENDIX IV PROOF OF LEMMA 4

*Proof:* The relationship between moments and cumulants gives [22]

$$\sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} = \sum_{i,j=1}^n [\kappa(\varepsilon_i^2 \varepsilon_j^2) + \kappa(\varepsilon_i^2) \kappa(\varepsilon_j^2) + 2\kappa^2(\varepsilon_i \varepsilon_j)] \quad (96)$$

where  $\kappa(\cdot)$  denotes the cumulants. It has been shown that [23]

$$\kappa(\varepsilon_i^2 \varepsilon_j^2) \sim O((n+2)^{-3}) \quad \text{as } n \text{ large}. \quad (97)$$

Hence, it is sufficient to study the orders of the summations of the remaining two terms, which are of order  $O((n+2)^{-2})$  [23]. It follows that  $\kappa(\varepsilon_i^2) = \sigma_i^2$ ,  $\kappa(\varepsilon_j^2) = \sigma_j^2$ , and  $\kappa(\varepsilon_i \varepsilon_j) = \mathcal{C}(\varepsilon_i, \varepsilon_j) = \sigma_{i,j}$  (say) [13]. Substituting these identities into (96) and ignoring the terms  $\kappa(\varepsilon_i^2 \varepsilon_j^2)$ , we have

$$\sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} \simeq \sum_{i,j=1}^n (\sigma_i^2 \sigma_j^2 + 2\sigma_{i,j}^2). \quad (98)$$

Applying the Cauchy-Schwarz inequality yields  $\sigma_{i,j}^2 \leq \sigma_i^2 \sigma_j^2$ . Recalling from Lemma 3 that  $\sum \sigma_i^2 = O(\log \log n)$ , we then have

$$\sum_{i,j=1}^n \sigma_{i,j}^{(2,2)} \leq 3 \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n \sigma_j^2 = O((\log \log n)^2) \quad (99)$$

thus completes the proof. ■

#### REFERENCES

- [1] R. A. Fisher, *Statistical Methods, Experimental Design, and Scientific Inference*. New York: Oxford Univ. Press, 1990.
- [2] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, pp. 507–521, 1915.

- [3] R. A. Fisher, "On the 'probable error' of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, pp. 3–32, 1921.
- [4] E. Fieller, H. Hartley, and E. Pearson, "Test for rank correlation coefficients. I," *Biometrika*, vol. 44, pp. 470–481, 1957.
- [5] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, 5th ed. New York: Oxford Univ. Press, 1990.
- [6] D. D. Mari and S. Kotz, *Correlation and Dependence*. London, U.K.: Imperial College Press, 2001.
- [7] W. Xu, C. Chang, Y. S. Hung, S. K. Kwan, and P. C. W. Fung, "Order statistic correlation coefficient and its application to association measurement of biosignals," in *Proc. ICASSP'06*, 2006, vol. 2, pp. II-1068–II-1071.
- [8] W. Xu, C. Chang, Y. S. Hung, S. K. Kwan, and P. C. W. Fung, "Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5552–5563, Dec. 2007.
- [9] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, 3rd ed. New York: Marcel Dekker, 1992.
- [10] H. David and H. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.
- [11] N. Balakrishnan and C. R. Rao, *Order Statistics: Applications*. New York: Elsevier, 1998, vol. 17, Handbook of Statistics.
- [12] N. Balakrishnan and C. R. Rao, *Order Statistics: Theory & Methods*. New York: Elsevier, 1998, vol. 16, Handbook of Statistics.
- [13] A. Stuart, J. K. Ord, and S. F. Arnold, *Kendall's Advanced Theory of Statistics: Volume 1 Distribution Theory*, 6th ed. London, U.K.: Edward Arnold, 1994.
- [14] D. S. Mitrinovic, J. E. Pecaric, and A. M. Fink, *Classical and New Inequalities in Analysis*. Dordrecht, The Netherlands: Kluwer Academic, 1993.
- [15] O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques For Use in Statistics*. London, U.K.: Chapman & Hall, 1989.
- [16] A. W. Vaart, *Asymptotic Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [17] H. Hotelling, "New light on the correlation coefficient and its transforms," *J. R. Statist. Soc. Ser. B (Methodological)*, vol. 15, pp. 193–232, 1953.
- [18] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics: Volume 2 Classical Inference and Relationship*, 5th ed. London, U.K.: Edward Arnold, 1991.
- [19] E. Schechtman and S. Yitzhaki, "A measure of association base on Gini's mean difference," *Commun. Statist.-Theor. Meth.*, vol. 16, no. 1, pp. 207–231, 1987.
- [20] W. Xu, C. Chang, and Y. S. Hung, "The asymptotic distribution of sample Gini correlation in normal cases," *J. Multivariate Anal.*, 2008, submitted for publication.
- [21] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, pp. 32–52, 1928.
- [22] R. A. Fisher, "Moments and product moments of sampling distributions," *Proc. London Math. Soc.*, vol. 30, pp. 199–238, 1929.
- [23] F. David and N. Johnson, "Statistical treatment of censored data. Part I: Fundamental formulae," *Biometrika*, vol. 41, pp. 228–240, 1954.



**Weichao Xu** (M'06) received the B.Eng. and M.Eng. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1993 and 1996, respectively, and the Ph.D. degree in biomedical engineering from the University of Hong Kong, Hong Kong, in 2002.

Since 2003, he has been a Research Associate with the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include the areas of mathematical statistics, pattern recognition, digital signal processing, and

applications.



**Chunqi Chang** (M'06) received the B.Sc. and M.Sc. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 1992 and 1995, respectively, and the Ph.D. degree in biomedical engineering from the University of Hong Kong, Hong Kong, in 2001.

He is currently a Research Fellow with the Department of Electrical and Electronic Engineering, University of Hong Kong. His main research interests include statistical signal processing theories and methods and their applications to biomedical engineering and computational molecular biology.



**Y. S. Hung** (M'88–SM'02) received the B.Sc. (Eng.) degree in electrical engineering and the B.Sc. degree in mathematics from the University of Hong Kong, Hong Kong, and the M.Phil. and Ph.D. degrees from the University of Cambridge, Cambridge, U.K.

He was a Research Associate with the University of Cambridge and a Lecturer with the University of Surrey, Surrey, U.K. In 1989, he joined the University of Hong Kong, where he is currently an Associate Professor and the Head of the department.

His research interests include robust control systems theory, robotics, computer vision, and biomedical engineering.

Dr. Hung was a recipient of the Best Teaching Award in 1991 from the Hong Kong University Students' Union. He is a chartered engineer and a fellow of IET and HKIE.



**Peter Chin Wan Fung** received the B.Sc. (Phys.) degree, the B.Sc. degree (special honors) in radio astronomy, and the Ph.D. degree in radio astronomy from the University of Tasmania, Hobart, Australia.

In 1999, he became the first Chair Professor of Medical Physics with the Department of Medicine, University of Hong Kong, Hong Kong. He is also an Honorary Professor in the Department of Psychiatry and the Department of Electrical and Electronic Engineering, University of Hong Kong, where he works on multidisciplinary projects. He has been

with the University of Montreal, Montreal, QC, Canada, and Stanford University, Stanford, CA. During the early 1970s, he joined the University of Hong Kong and became the Personal (Chair) Professor of Physics in 1984 and the Director of the Centre for Materials Science in 1992. He has published around 280 articles in international reviewed journals. His current research interests include the areas of biophysics, medicine, and signal processing.